Bounds on the distance between a unital quantum channel and the convex hull of unitary channels, with applications to the asymptotic quantum Birkhoff conjecture

Nengkun Yu*

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
and Centre for Quantum Computation and Intelligent Systems (QCIS),
Faculty of Engineering and Information Technology,
University of Technology, Sydney, NSW 2007, Australia

Runyao Duan†

Centre for Quantum Computation and Intelligent Systems (QCIS),
Faculty of Engineering and Information Technology,
University of Technology, Sydney, NSW 2007, Australia
and State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Quanhua Xu‡

School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
and Laboratoire de Mathématiques, Université de Franche-Comté, 25030 Besançon cedex, France

Motivated by the recent resolution of Asymptotic Quantum Birkhoff Conjecture (AQBC), we attempt to estimate the distance between a given unital quantum channel and the convex hull of unitary channels. We provide two lower bounds on this distance by employing techniques from quantum information and operator algebras, respectively. We then show how to apply these results to construct some explicit counterexamples to AQBC. We also point out an interesting connection between the Grothendieck's inequality and AQBC.

I. INTRODUCTION

Suppose we are given a quantum system with a d-dimensional Hilbert space $\mathcal{H}_d$, and the state (or density operator) of the system is given by a trace one positive operator ρ from the linear operator space $\mathrm{L}(\mathcal{H}_d)$. Quantum channels, or trace-preserving completely positive maps, are all possible deterministic quantum operations one can perform over the system [1, 2]. Let Φ be such a quantum channel over $\mathrm{L}(\mathcal{H}_d)$ with Kraus operator sum representation $\Phi = \sum_k E_k \cdot E_k^\dagger$, and let $K(\Phi) = \mathrm{span}\{E_k\}$ be its Kraus operator space. The convex hull of unitary channels (noiseless channels) on $\mathrm{L}(\mathcal{H}_d)$ is given by $\mathrm{Conv}(\mathbb{U}(\mathcal{H}_d))$. So any $\Psi \in \mathrm{Conv}(\mathbb{U}(\mathcal{H}_d))$ can be written as a mixture (convex combination) of unitary channels. (The number of unitary channels in the mixture can be made finite due to the Carathéodory's theorem on convex hull). The mixture of unitary channels plays a special role in environment-assisted quantum communication model. Actually, these channels can be made noiseless for quantum information transmission with the help of a friendly environment even in one-shot case. Furthermore, it turns out that these channels are the only quantum channels having this desirable property [3]. Surprisingly, if arbitrarily large number of uses of the channels are allowed, unital quantum channels, those channels Φ with identity operator a fixed point, say $\Phi(I) = I$, can also achieve maximum capacity and act exactly like noiseless channel [4].

Clearly, any mixture of unitary channels remains unital. An interesting question is to ask whether one can reverse this procedure, i.e., decomposing any unital quantum channel $\Phi \in \mathrm{T}(\mathcal{H}_d)$ into a mixture of unitary channels from $\mathbb{U}(\mathcal{H}_d)$. This was called "quantum Birkhoff conjecture" (QBC), originated from Birkhoff's celebrated characterization of the extreme points of doubly stochastic matrices. Unfortunately, this conjecture is only true for $d \leq 2$, and counterexamples exist whenever $d \geq 3$ [5–7]. This suggests the following quantity to measure the distance between Φ and the convex hull of unitary channels.

$$D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H}))) = \inf\{D(\Phi, \Psi) : \Psi \in \mathrm{Conv}(\mathbb{U}(\mathcal{H}))\},$$

where $D(\Phi, \Psi)$ will be given by the diamond norm of $\Phi - \Psi$. Since $\mathrm{Conv}(\mathbb{U}(\mathcal{H}))$ is a compact convex set, "inf" in the above equation can be replaced by "min".

*Electronic address: nengkunyu@gmail.com
†Electronic address: runyao.duan@uts.edu.au
‡Electronic address: qxu@univ-fcomte.fr

Motivated by some results in about the environment-assisted quantum capacity and in an attempt to remedy the conjecture in certain way, Smolin, Verstraete, and Winter proposed the following

Conjecure 1. *(Asymptotic Quantum Birkhoff Conjecture [4]) Let $\Phi \in \mathrm{T}(\mathcal{H})$ be a unital channel, then $\Phi^{\otimes n}$ can be approximated by a mixture of unitary channels from $\mathbb{U}(\mathcal{H}^{\otimes n})$ with arbitrary precision. That is*

$$\lim_{n \to \infty} D(\Phi^{\otimes n}, \mathrm{Conv}(\mathbb{U}(\mathcal{H}^{\otimes n}))) = 0.$$

This revised conjecture seems highly reasonable as one could naturally expect that many copies of a unital channel will be better approximated by a mixture of unitary channels on a higher-dimensional space. If this is true, it will provide a very satisfactory interpretation to the following result: The environment-assisted quantum capacity of any unital channel over $\mathrm{L}(\mathcal{H}_d)$ is given by $\log_2 d$ qubits, the maximum capacity one can achieve under this model. A much more deep consequence is that the structure of unital channels will be greatly simplified. Due to its significance, the asymptotic quantum Birkhoff conjecture was listed as one of major open problems in quantum information theory [8].

Some supporting evidences were obtained in Ref. [10], where Mendl and Wolf presented a unital channel Φ such that $\Phi^{\otimes 2}$ is a mixture of unitary channels although Φ itself is not. Furthermore, they showed that it is possible that the tensor of Φ and a constant unital channel (a completely depolarizing channel that maps every state into the completely mixed state I/d) may become a mixture of unitary channels. One may naturally conjecture these properties might be true for any unital quantum channels.

Recently Haagerup and Musat disproved this asymptotic version by exhibiting a class of so-called non-factorizable maps as counterexamples [11]. Actually the results obtained in Ref. [11] shows that any such non-factorizable map Φ is a very strong counterexample to AQBC in the following sense:

$$D(\Phi \otimes \Psi, \mathcal{FM}(\mathrm{L}(\mathcal{H}_d \otimes \mathcal{H}_m))) \geq D(\Phi, \mathcal{FM}(\mathrm{L}(\mathcal{H}_d))),$$

where Ψ is any unital channel over $\mathrm{L}(\mathcal{H}_m)$, and $\mathcal{FM}(\mathrm{L}(\mathcal{H}_d))$ denotes the set of factorizable maps over $\mathrm{L}(\mathcal{H}_d)$. In other words, any non-factorizable map tensoring with a unital channel could not reduce the distance to the set of factorizable maps, which is a super-set of the convex hull of unitary channels. See also Shor's talk in Ref. [12] for an alternative approach to AQBC and an excellent discussion of the results in Ref. [11]. The interesting thing here is that all these counterexamples are non-factorizable maps, and it remained unknown whether any facterizable map would fulfill AQBC. This problem was signified in the arXiv version of Ref. [11] by establishing the following surprising connection: If all factorizable maps satisfy AQBC, then the Connes embedding problem has a positive answer.

Motivated by these progresses and in order to better understand the structure of unital channels, in this paper we are interested in estimating the trace distance between a unital quantum channel and the convex hull of unitary channels, say $D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H})))$. We find that this distance is interesting even from the perspective of quantum channel discrimination: Suppose we are given an unknown quantum channel, which is secretly chosen between Φ and some $\Psi \in \mathrm{Conv}(\mathbb{U}(\mathcal{H}))$ with equal probability $1/2$. Then due to the operational meaning of trace distance, we can conclude that the success probability of discrimination is at least $1/2 + 1/4D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H})))$, which is strictly larger than $1/2$ whenever Φ is not a mixture of unitary channels. Another purpose of this paper is to provide some relatively elementary and self-contained disproofs to AQBC. This is partially due to the fact that the elegant disproof of AQBC in Ref. [11] makes use of some basic properties of factorizable maps which cannot be easily appreciated by readers who do not have deep background in operator algebras.

In Section II we collect some preliminaries about super-operators and Schur channels. Then in Section III we explain in detail the operational meaning of trace distance. In Section IV we first provide a computable lower bound for $D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H})))$ when the Kraus operator space of Φ does not contain any unitary operator. This enables us to derive many counterexamples for AQBC, including some factorizable maps presented in Ref. [11]. It is worth pointing out that this proof only employs some basic techniques from quantum information theory. We believe that it may interest readers with quantum information background. In Section V we go further to study the class of Schur channels. In this special case, we are able to provide a lower bound and an upper bound for $D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H})))$. Roughly speaking, we show that up to a factor of $1/2$, any Schur channel can be approximated by a mixture of diagonal unitary channels, and the later has a simpler structure. As a direct application, we obtain a new proof of the fact that any Schur channel that does not satisfy the QBC will automatically violate the AQBC. Our proof for this part has employed some powerful tools from operator algebras. In Section VI we present two explicit examples of Schur channels to demonstrate the utility of our results: the first example has only two Kraus operators and is a non-factorizable map, and the second one is a factorizable map. As another interesting application, in Section VII we point out a connection between AQBC and Grothendieck's inequality in the metric theory of tensor products.

Remarks on related results: After we obtained the results in Section IV, and were working on the proof of the Theorem 3 in Section V, the second author R.D. happened to learn from Prof. M. B. Ruskai that Haagerup and Musat had made further progress on the connection between Schur channels and AQBC. Namely, they obtained Theorem 5 and thus showed that any Schur channel that violates QBC (including some factorizable maps) should also be a counterexample to AQBC [13].

They also provided a modified version of the connection between factorizable maps satisfying AQBC and Connes embedding problem. The proof of Theorem 3 has employed some similar techniques in [13].

II. PRELIMINARIES

We will use symbols $\mathcal{H}$, $\mathcal{K}$, etc to represent finite dimensional Hilbert spaces over complex numbers. A d-dimensional Hilbert space $\mathcal{H}$, which is essentially the same as $\mathcal{C}^d$, will be explicitly represented as $\mathcal{H}_d$ whenever the dimension matters. $\mathrm{L}(\mathcal{H},\mathcal{K})$ denotes the set of linear operators (or mappings) from $\mathcal{H}$ to $\mathcal{K}$, and $\mathrm{L}(\mathcal{H})$ is shorthand for $\mathrm{L}(\mathcal{H},\mathcal{H})$. For any $X \in L(\mathcal{H})$, $X^\dagger \in L(\mathcal{H})$ denotes the adjoint operator (or complex conjugate) of X. X is Hermitian (or self-adjoint) if $X^\dagger = X$. $U \in \mathrm{L}(\mathcal{H})$ is said to be unitary if $U^\dagger U = I_\mathcal{H}$. We denote the set of unitary operators on $\mathcal{H}$ by $\mathrm{U}(\mathcal{H})$. $X \in \mathrm{L}(\mathcal{H})$ is (semi-definite) positive, write $X \geq 0$, if the quadratic form $\langle\psi|X|\psi\rangle \geq 0$ for any $|\psi\rangle \in \mathcal{H}$. In particular, X is said to be a density operator (or a quantum state) if X is positive and with trace one. $\mathrm{T}(\mathcal{H},\mathcal{K})$ is the set of linear mappings from $\mathrm{L}(\mathcal{H})$ to $\mathrm{L}(\mathcal{K})$. Again, $\mathrm{T}(\mathcal{H})$ is shorthand for $\mathrm{T}(\mathcal{H},\mathcal{H})$. Elements in $\mathrm{T}(\mathcal{H},\mathcal{K})$ are normally called super-operators. Note that $\mathrm{L}(\mathcal{H})$ is a Hilbert space with the standard Hilbert-Schmidt inner product $< A, B > = \mathrm{Tr}(A^\dagger B)$. Then the adjoint operator of $\Phi \in \mathrm{T}(\mathcal{H},\mathcal{K})$ is defined as the unique super-operator $\Phi^\dagger \in \mathrm{T}(\mathcal{K},\mathcal{H})$ such that

$$< Y, \Phi(X) > = < \Phi^\dagger(Y), X >, \quad \forall X \in \mathrm{L}(\mathcal{H}), Y \in \mathrm{L}(\mathcal{K}).$$

A super-operator $\Phi \in \mathrm{T}(\mathcal{H},\mathcal{K})$ is said to be positive if it preserves the positivity, say, $\Phi(X) \geq 0$ whenever $X \geq 0$. Φ is said to be a quantum channel if it satisfies: i) (trace-preserving) $\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}(X)$ for any $X \in \mathrm{L}(\mathcal{H})$, and ii)(completely positive) for any $n \geq 1$, the induced super-operator $\Phi_n = \Phi \otimes I_{\mathrm{L}(\mathcal{H}_n)} \in \mathrm{T}(\mathcal{H} \otimes \mathcal{H}_n, \mathcal{K} \otimes \mathcal{H}_n)$ is positive, where $I_{\mathrm{L}(\mathcal{H}_n)}$ is the identity super-operator on $\mathrm{L}(\mathcal{H}_n)$. We call Φ a quantum unital channel if it further satisfies: iii) (unital condition) $\Phi(I_\mathcal{H}) = I_\mathcal{K}$. Any unitary operator $U \in \mathrm{U}(\mathcal{H})$ induces a unitary quantum channel $\mathcal{U} \in \mathrm{T}(\mathcal{H})$ in the following way: $\mathcal{U}(X) = UXU^\dagger$. The class of unitary channels on $\mathrm{L}(\mathcal{H})$ will be denoted as $\mathbb{U}(\mathcal{H})$.

Any super-operator $\Phi \in \mathrm{T}(\mathcal{H},\mathcal{K})$ can be represented by a pair of linear operators $A, B \in \mathrm{L}(\mathcal{H}, \mathcal{K} \otimes \mathcal{Z})$ such that

$$\Phi(X) = \mathrm{Tr}_\mathcal{Z} A X B^\dagger, \quad X \in \mathrm{L}(\mathcal{H}), \tag{1}$$

where $\mathcal{Z}$ is an auxiliary Hilbert space with $\dim(\mathcal{Z}) \leq \dim(\mathcal{H})\dim(\mathcal{K})$, and $\mathrm{Tr}_\mathcal{Z}$ represents the partial trace over $\mathcal{Z}$. For the special case of quantum channels, the above form can be greatly simplified. Actually, in Eq. (1) we can choose $A = B = V \in \mathrm{L}(\mathcal{H}, \mathcal{K} \otimes \mathcal{Z})$ for some isometry V and obtain the following well-known Stinespring unitary embedding representation of a quantum channel:

$$\Phi(X) = \mathrm{Tr}_\mathcal{Z} V X V^\dagger, \quad V^\dagger V = I_\mathcal{H}. \tag{2}$$

If we specify an orthonormal basis $\{|k_\mathcal{Z}\rangle\}$ of $\mathcal{Z}$, we can rewrite Φ in Eq (1) into the following form:

$$\Phi(X) = \sum_{k=1}^{\dim(\mathcal{Z})} A_k X B_k^\dagger, \tag{3}$$

where $A_k = \langle k_\mathcal{Z}|A$ and $B_k = \langle k_\mathcal{Z}|B$ are linear operators in $\mathrm{L}(\mathcal{H},\mathcal{K})$. Similarly, when Φ is a quantum channel, we can choose $A_k = B_k = \langle k_\mathcal{Z}|V$ so that

$$\Phi(X) = \sum_{k=1}^{\dim(\mathcal{Z})} A_k X A_k^\dagger, \quad \sum_k A_k^\dagger A_k = I_\mathcal{H}, \tag{4}$$

which is the famous Kraus operator sum representation of a quantum channel [1].

Now we tend to introduce norms of super-operators in $\mathrm{T}(\mathcal{H},\mathcal{K})$ based on the norms of linear operators. We refer to Refs. [14, 15] for some detailed discussion on norms of super-operators and how to compute them using semidefinite programming techniques. We will briefly review some basic results for later use. For any $X \in \mathrm{L}(\mathcal{H}_d)$ and $p \geq 1$, the p-th norm of X is given by

$$||X||_p = (\mathrm{Tr}|X|^p)^{\frac{1}{p}},$$

where $|X| = \sqrt{X^\dagger X}$. The trace and the operator norms of X are special cases of $p = 1$ and $p \to \infty$, respectively,

$$||X||_1 = \mathrm{Tr}|X|, \; ||X||_\infty = \max_{\langle\psi|\psi\rangle=1} ||X|\psi\rangle||.$$

The trace norm and the operator norm of a super-operator $\Phi \in \mathrm{T}(\mathcal{H},\mathcal{K})$ are given respectively as follows:

$$||\Phi||_1 = \sup_{||X||_1 \leq 1} ||\Phi(X)||_1, \; ||\Phi||_\infty = \sup_{||X||_\infty \leq 1} ||\Phi(X)||_\infty.$$

In the above equation we can replace "sup" with "max" when only finite dimensional Hilbert spaces are involved. The completely bounded trace norm (or diamond norm) and operator norm (simply completely bounded norm) are given respectively as follows:

$$||\Phi||_\diamond = \sup_{n \geq 1} ||\Phi \otimes I_{\mathrm{L}(\mathcal{H}_n)}||_1, \; ||\Phi||_{\mathrm{cb}} = \sup_{n \geq 1} ||\Phi \otimes I_{\mathrm{L}(\mathcal{H}_n)}||_\infty.$$

Proposition 1. *For any $\Phi \in \mathrm{T}(\mathcal{H},\mathcal{K})$, the diamond norm and the completely bounded norm satisfy the following properties:*

- *i) The dimension of the auxiliary system to achieve the norms can be restricted to that of $\mathcal{H}$, $||\Phi||_\diamond = ||\Phi \otimes I_{\mathrm{L}(\mathcal{H})}||_1$ and $||\Phi||_{\mathrm{cb}} = ||\Phi \otimes I_{\mathrm{L}(\mathcal{H})}||_\infty$.*

- *ii) The following duality relation holds for Φ and $\Phi^\dagger$, $||\Phi||_1 = ||\Phi^\dagger||_\infty$ and $||\Phi||_\diamond = ||\Phi^\dagger||_{\mathrm{cb}}$.*

- *iii) If Φ is completely positive, then $||\Phi||_\diamond = ||\Phi||_1$ and $||\Phi||_{\mathrm{cb}} = ||\Phi||_\infty = ||\Phi(I_\mathcal{H})||_\infty$.*

The norms defined above enable us to introduce distance between quantum states and quantum channels. The trace distance between two quantum density operators ρ and σ in $L(\mathcal{H})$ is given by

$$D(\rho, \sigma) = ||\rho - \sigma||_1.$$

In the following discussion we also need the fidelity between ρ and σ,

$$F(\rho, \sigma) = \text{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}.$$

The so-called Uhlmann theorem makes the meaning of fidelity more transparent:

$$F(\rho, \sigma) = \max_{|\psi\rangle, |\phi\rangle} |\langle\psi|\phi\rangle|,$$

where $|\psi\rangle, |\phi\rangle \in \mathcal{H} \otimes \mathcal{K}$ range over all purifications of ρ and σ, respectively, say $\text{Tr}_{\mathcal{K}}|\psi\rangle\langle\psi| = \rho$ and $\text{Tr}_{\mathcal{K}}|\phi\rangle\langle\phi| = \sigma$. Most notably, the above equation remains true even when one of $|\psi\rangle$ or $|\phi\rangle$ is fixed. This fact plays a crucial role in our later discussion. Trace distance and fidelity are equivalent in characterizing the distance between two states in the following sense:

$$2(1 - F(\rho, \sigma)) \leq D(\rho, \sigma) \leq 2\sqrt{1 - F^2(\rho, \sigma)}.$$

Following the same idea, we can define the trace distance between two quantum channels Φ and Ψ via the following way:

$$D(\Phi, \Psi) = ||\Phi - \Psi||_\diamond.$$

Let us now introduce a special class of super-operators. For any $S \in L(\mathcal{H}_d)$, we can define a super-operator Φ_S via the following way:

$$\Phi_S(X) = S \circ X, \ \forall X \in L(\mathcal{H}_d),$$

where $S \circ X = [s_{kj}x_{kj}]$ is the entry-wise product or Hadamard product. (Here we assume that we have specified an orthonormal basis $\{|k\rangle : k = 1, \cdots, d\}$ for $\mathcal{H}_d$. Thus any linear operator from $L(\mathcal{H}_d)$ is expressed as a matrix under the standard matrix basis $\{|k\rangle\langle j|\}$. For instance, $S = \sum_{k,j} s_{kj}|k\rangle\langle j|$). Such Φ_S is called Schur multiplier induced by S. Schur multipliers have been extensively studied in the literatures of operator algebras. We refer to Chapters 3 and 8 of Ref. [9] for some highly accessible introductions, and Ref. [11] for recent advances. For later use, some basic properties of Schur multipliers are listed as follows:

Proposition 2. *Let $S \in L(\mathcal{H}_d)$. Then Φ_S satisfies the following:*

- *i) $\Phi_S = \sum_{k=1}^d A_k \cdot B_k^\dagger$, where all A_k, B_k are diagonal matrices;*

- *ii) Φ_S is positivity-preserving iff S is positive;*

- *iii) Φ_S is completely positive iff S is positive;*

- *iv) Φ_S is trace-preserving if $s_{kk} = 1$ for $k = 1, \cdots, d$;*

- *v) Φ_S is unital iff $s_{kk} = 1$ for $k = 1, \cdots, d$.*

Proof: *iv)* and *v)* follow directly by evaluating $\text{Tr}(\Phi_S(|k\rangle\langle j|)) = s_{kj}\delta_{kj}$. We shall see that *ii)* and *iii)* are simple corollaries of *i)*. So we first prove *i)*. In fact, let A and B be any two $d \times d$ matrices such that $S = AB^\dagger$. We may assume $A = [a_1, \cdots, a_d]$ and $B = [b_1, \cdots, b_d]$, where a_k and b_k are all d-dimensional column vectors. Then $S = \sum_k a_k b_k^\dagger$. Set $A_k = \text{Diag}(a_k)$, $B_k = \text{Diag}(b_k)$. That is, A_k and B_k are diagonal matrices with diagonals a_k and b_k, respectively. By some routine calculations we directly verify that $\Phi_S(X) = \sum_{k=1}^d A_k X B_k^\dagger$ for any $X \in L(\mathcal{H}_d)$. In particular, when S is positive we can write $S = AA^\dagger$ for some $A \in L(\mathcal{H}_d)$. Hence we can choose $A_k = B_k$ in this special case. That proves both the positivity and completely positivity of Φ_S. Conversely, if Φ_S is positive. Then by choosing $|e\rangle = \sum_{k=1}^d |k\rangle$, we have $\Phi_S(|e\rangle\langle e|) = S$ is positive. $\square$

The following proposition gives another fundamental property of Schur multiplier. Relevant discussions can be found in Page 110 of Ref. [9].

Proposition 3. *For any Schur multiplier Φ, the diamond norm, the trace norm, completely bounded norm and operator norm all coincide, that is, $||\Phi||_\diamond = ||\Phi||_1 = ||\Phi||_{\text{cb}} = ||\Phi||_\infty$.*

So a Schur multiplier Φ_S is a quantum channel iff S is positive and with all diagonal entries one. In particular, whenever Φ_S is a quantum channel, it is also unital. We shall denote

$$S(\mathcal{H}_d) = \{\Phi_S : S \in L(\mathcal{H}_d), S \geq 0, s_{kk} = 1, 1 \leq k \leq d\},$$

and call the elements from $S(\mathcal{H}_d)$ (or simply S_d) Schur channels. Note that the difference of two Schur multipliers is still a Schur multiplier. Applying Proposition 3, we obtain an immediate consequence that auxiliary systems are not required to distinguish between two Schur channels.

III. OPERATIONAL INTERPRETATION OF TRACE DISTANCE

We have introduced trace distance between quantum states and quantum channels, and will study the trace distance between a unital quantum channel and the convex hull of unitary channels in greater detail. Before we proceed, we need justify the importance of this measure from the perspective of quantum information. In one word, the trace distance characterizes some sort of stochastic distinguishability of quantum states and quantum channels. Actually, the trace distance naturally occurs when we study the following state discrimination problem. Suppose we are given an unknown quantum system whose state is secretly prepared in one of ρ_0 and

ρ_1, with equal priori probability 1/2. The task here is to determine the identity of the system with a success probability as high as possible. To do so we need apply a two-outcome quantum measurement $\{E_0, E_1\}$ to the system, and to maximize the success probability of discrimination, i.e.,

$$P_{\text{succ}}(\rho_0, \rho_1) = \max_{\{E_0, E_1\}} \frac{1}{2}(\text{Tr}\rho_0 E_0 + \text{Tr}\rho_1 E_1),$$

where $E_i \geq 0$ and $E_0 + E_1 = I$. By some simple algebraic manipulations, one can verify that the optimal success probability of discrimination is given by [16]

$$P_{\text{succ}}(\rho_0, \rho_1) = \frac{1}{2} + \frac{1}{4}D(\rho_0, \rho_1).$$

Thus a larger trace distance between ρ_0 and ρ_1 implies a higher success probability of discrimination. This interpretation can be extended to compact convex sets of density operators. Let A_0 and A_1 be two compact convex sets of density operators. The trace distance between A_0 and A_1 is given by

$$D(A_0, A_1) = \min\{D(\rho_0, \rho_1) : \rho_i \in A_i, i = 0, 1\}.$$

Then the optimal discrimination probability between A_0 and A_1 is given as

$$P_{\text{succ}}(A_0, A_1) = \frac{1}{2} + \frac{1}{4}D(A_0, A_1). \qquad (5)$$

The above formula indicates that we can operationally distinguish between two compact convex sets of density operators by performing a universal quantum measurement, and the success probability of discrimination is completely characterized by the trace distance between A_0 and A_1. The most surprising thing here is that the quantum measurement we perform does not depend on the exact form of the unknown state except the assumption that it is from one of A_0 and A_1.

It seems that Eq. (5) was first obtained by Gutoski and Watrous in Ref. [18] by using the convex set separation theorem. Jain provided a different way based on the minimax theorem [19]. For completeness, we will outline the later approach as follows. Let $\{E_0, E_1\}$ be the quantum measurement we need perform, and ρ_0 and ρ_1 be two states from A_0 and A_1, respectively. Then the optimal success probability is given by

$$P_{\text{succ}}(A_0, A_1) = \max_{\{E_i\}} \min_{\rho_i \in A_i} \frac{1}{2}(\text{Tr}\rho_0 E_0 + \text{Tr}\rho_1 E_1).$$

The crucial point here is that we first take "min" over all possible pair of states ρ_0 and ρ_1 according to a fixed measurement $\{E_0, E_1\}$, and then take "max" over all possible measurements to maximize the success probability of discrimination. Noticing that the objective function is linear in (E_0, E_1) and (ρ_0, ρ_1) when one of them is fixed, and all involving sets are compact convex, we can apply

appropriate form of Sion's minimax theorem [17] to exchange the order of "max" and "min", and obtain Eq. (5) immediately.

Now we try to generalize the above result to the case of quantum channels. The simplest case is to distinguish between two quantum channels $\Phi_0, \Phi_1 \in \text{T}(\mathcal{H}, \mathcal{K})$. The basic strategy here is to choose an input state $\rho \in \text{L}(\mathcal{H}' \otimes \mathcal{H})$, and then to distinguish between the respective output states $I_{\text{L}(\mathcal{H}')} \otimes \Phi_i(\rho)$, where $\mathcal{H}'$ is a finite-dimensional auxiliary state space. We have

$$D(\Phi_0, \Phi_1; \rho) = D((I_{\text{L}(\mathcal{H}')} \otimes \Phi_0)(\rho), (I_{\text{L}(\mathcal{H}')} \otimes \Phi_1)(\rho)).$$

To achieve the maximum success probability, we need take "sup" over all possible input states, and have

$$D(\Phi_0, \Phi_1) = \sup_{\rho} D(\Phi_0, \Phi_1; \rho).$$

One can readily verify that the RHS of the above equation gives us the diamond norm $\|\Phi_0 - \Phi_1\|_\diamond$, and ρ can be restricted to density operators on $\mathcal{H} \otimes \mathcal{H}$ (thus "sup" can be replaced as "max"). To generalize the trace distance to compact convex sets of quantum channels, we first need the trace distance with input state ρ as follows:

$$\widetilde{D}(C_0, C_1; \rho) = \min_{\Phi_i \in C_i} D(\Phi_0, \Phi_1; \rho).$$

Then the final resulting operational trace distance between C_0 and C_1 is given by

$$\widetilde{D}(C_0, C_1) = \sup_{\rho} \widetilde{D}(C_0, C_1; \rho) = \sup_{\rho} \min_{\Phi_i \in C_i} D(\Phi_0, \Phi_1; \rho), \qquad (6)$$

where ρ ranges over all possible bipartite density operators on $\mathcal{H}' \otimes \mathcal{H}$, and it is not clear whether we can replace "sup" with "max" as the dimension of $\mathcal{H}'$ may be arbitrarily large. The optimal success probability of discrimination between C_0 and C_1 is given by

$$P_{\text{succ}}(C_0, C_1) = \frac{1}{2} + \frac{1}{4}\widetilde{D}(C_0, C_1).$$

Interestingly, the (ordinary) trace distance between C_0 and C_1 is given by

$$D(C_0, C_1) = \min_{\Phi_i \in C_i} D(\Phi_0, \Phi_1) = \min_{\Phi_i \in C_i} \max_{\rho} D(\Phi_0, \Phi_1; \rho).$$

The major difference between $D(C_0, C_1)$ and $\widetilde{D}(C_0, C_1)$ is that the orders of "max" ("sup") and "min" has been reversed. It is not obvious that whether the orders of "min" and "max" ("sup") are exchangeable or not as it is unclear whether the objective function $D(\Phi_0, \Phi_1; \rho)$ satisfies the requirements of minimax theorem. Consequently, it seems not clear whether $\widetilde{D}(C_0, C_1)$ is the same as $D(C_0, C_1)$. Nevertheless, we still have

$$\widetilde{D}(C_0, C_1; \rho) \leq \widetilde{D}(C_0, C_1) \leq D(C_0, C_1).$$

In particular, we have the following simple property.

Property 1. *Let* $C_0, C_1 \subseteq T(\mathcal{H}, \mathcal{K})$ *be two compact convex sets of quantum channels, and let* ρ *be a bipartite pure entangled state over* $\mathcal{H} \otimes \mathcal{H}$ *with full Schmidt rank. Then the following are equivalent:*

> *i).* $C_0 \cap C_1 = \emptyset$;
>
> *ii).* $D(C_0, C_1) > 0$;
>
> *iii).* $\widetilde{D}(C_0, C_1) > 0$; *and*
>
> *iv).* $\widetilde{D}(C_0, C_1; \rho) > 0$.

Proof: We only need to establish the equivalence between i) and iv). By definition, iv) means that we can distinguish between C_0 and C_1 using ρ as an input. This immediately implies that C_0 and C_1 should be disjoint. In other words, i) should hold. The direction that i)$\Rightarrow$ iv) is a little bit tricky, and the key here is to apply a generalized form of Choi isomorphism [2] between super-operators and bipartite linear operators. By contradiction, assume that C_0 and C_1 are disjoint but $\widetilde{D}(C_0, C_1; \rho) = 0$. It follows from the definition that there exist $\Phi_0 \in C_0$ and $\Phi_1 \in C_1$ such that

$$D((I_{L(\mathcal{H})} \otimes \Phi_0)(\rho), (I_{L(\mathcal{H})} \otimes \Phi_1)(\rho)) = 0.$$

Equivalently, we have

$$(I_{L(\mathcal{H})} \otimes \Phi_0)(\rho) = (I_{L(\mathcal{H})} \otimes \Phi_1)(\rho). \tag{7}$$

Noticing that ρ is a bipartite pure state with full Schmidt rank, we have the following generalized Choi-isomorphism:

$$J : \Phi \mapsto (I_{L(\mathcal{H})} \otimes \Phi)(\rho).$$

(The standard Choi-isomorphism is to choose ρ as the maximally entangled state $|\Omega\rangle = 1/\sqrt{d} \sum_{k=1}^{d} |k\rangle |k\rangle$). Applying this isomorphism, we deduce from Eq. (7) that $\Phi_0 = \Phi_1$. This contradicts the assumption $C_0 \cap C_1 = \emptyset$. $\square$

So whenever two compact convex sets of quantum channels are disjoint, we can operationally distinguish between them with a success probability strictly larger than $\frac{1}{2}$, and any bipartite pure state with full Schmidt rank can be used as input.

The really interesting thing here is that the equality of $\widetilde{D}(C_0, C_1) = D(C_0, C_1)$ does hold. The key to this is the application of Sion's minimax theorem and the following semi-definite programming characterization of the diamond norm recently discovered by Watrous [20].

Lemma 1. *(Watrous [15]) For any super-operator* $\Phi = \Phi_0 - \Phi_1$ *such that* Φ_0 *and* Φ_1 *are quantum channels in* $T(\mathcal{H}, \mathcal{K})$, *we have the following*

$$||\Phi||_\diamond = \max 2\text{Tr}\rho_\Phi X, \quad X \leq I \otimes \rho, \text{Tr}\rho = 1, \rho \geq 0, X \geq 0,$$

where $\rho_\Phi = (\Phi \otimes I_{\mathcal{H}'})(|\alpha\rangle\langle\alpha|)$ *is the Choi operator of* Φ, $|\alpha\rangle = \sum_{k=1}^{d} |k\rangle \otimes |k\rangle = \sqrt{d}|\Omega\rangle$ *is the unnormalized maximally entangled state over* $\mathcal{H} \otimes \mathcal{H}'$, *and* $\mathcal{H}'$ *is an isomorphic copy of* $\mathcal{H}$.

Now we can summarize the relation between $\widetilde{D}(C_0, C_1)$ and $D(C_0, C_1)$ as follows:

Theorem 1. *Let* C_0 *and* C_1 *be two compact convex sets of quantum channels in* $T(\mathcal{H}, \mathcal{K})$. *Then*

$$\widetilde{D}(C_0, C_1) = D(C_0, C_1).$$

Proof: Let us first denote

$$C = C_0 - C_1 = \{\Phi_0 - \Phi_1 : \Phi_0 \in C_0, \Phi_1 \in C_1\}.$$

Then C is a compact convex set, and completely determines $\widetilde{D}(C_0, C_1)$ and $D(C_0, C_1)$. We also write

$$R = \{(X, \rho) : 0 \leq X \leq I \otimes \rho, \rho \geq 0, \text{Tr}\rho = 1\}.$$

Clearly, R is also a compact convex set.

By Lemma 1, we can rewrite

$$D(C_0, C_1) = \min_{\Phi \in C} \max_{(X, \rho) \in R} 2\text{Tr}\rho_\Phi X.$$

Noticing that both C and R are compact convex sets, and the objective function $2\text{Tr}(\rho_\Phi X)$ is linear both in Φ and (X, ρ), by Sion's minimax theorem we can exchange the order of "max" and "min" as follows:

$$D(C_0, C_1) = \max_{(X, \rho) \in R} \min_{\Phi \in C} 2\text{Tr}\rho_\Phi X.$$

Now we proceed to prove $\widetilde{D}(C_0, C_1) = D(C_0, C_1)$. We only need to show $\widetilde{D}(C_0, C_1) \geq D(C_0, C_1)$ as the opposite direction is obvious according the definitions. By the above equation and Eq. (6), it suffices to show that for any $(X, \rho) \in R$ there is a density operator $\sigma \in L(\mathcal{H} \otimes \mathcal{H}')$ such that

$$||(\Phi \otimes I)(\sigma)||_1 \geq \text{Tr}\rho_\Phi X, \forall \Phi \in C.$$

Indeed, we can choose $\sigma = |u\rangle\langle u|$ to be the following bipartite pure state

$$|u\rangle = (I \otimes A)|\alpha\rangle \text{ and } A^\dagger A = \rho,$$

where $|\alpha\rangle$ is again the unnormalized maximally entangled state over $\mathcal{H} \otimes \mathcal{H}'$.

Note that we have the following well-known fact about the trace norm:

$$||Y||_1 = \max_{0 \leq P \leq I} 2\text{Tr}PY,$$

where Y is any traceless ($\text{Tr}Y = 0$) Hermitian operator. Applying the above fact to $(\Phi \otimes I)(\sigma)$, we have

$$||(\Phi \otimes I)(\sigma)||_1 = \max_P 2\text{Tr}P(I \otimes A)\rho_\Phi(I \otimes A^\dagger) = \max_Q 2\text{Tr}\rho_\Phi Q,$$

where $0 \leq P \leq I_{\mathcal{H} \otimes \mathcal{H}'}$ and $Q = (I \otimes A^\dagger)P(I \otimes A)$. Noticing that $0 \leq X \leq I \otimes \rho = I \otimes A^\dagger A$, we can easily find $0 \leq P' \leq I_{\mathcal{H} \otimes \mathcal{H}'}$ such that $X = Q' = (I \otimes A^\dagger)P'(I \otimes A)$ [21]. Thus we have

$$\max_Q 2\text{Tr}\rho_\Phi Q \geq 2\text{Tr}\rho_\Phi Q' = 2\text{Tr}\rho_\Phi X,$$

which completes the proof. $\qquad\square$

Remarks: After we finished the above proof, we were informed by Gutoski that in a recent work he generalized the results in Ref. [18] to the discrimination of two compact convex sets of quantum strategies, and obtained the results for the case of quantum channels as an immediate corollary [23]. It is interesting to note that his main proof technique is a separation theorem of compact convex sets from convex analysis, quite similar to that in Ref. [18]. Instead, here we employ a different method by using Sion's minimax theorem and semi-definite programming characterization of diamond norm, in a similar spirit of Ref. [19]. Hopefully, our proof may provide some new insight into this problem. Gutoski's paper, however, contains many other interesting results about the trace norms.

It is also worth noting that with minor changes the same technique in the above proof can be used to derive Lemma 1, as first shown by Watrous in Ref. [15].

All the above discussions are applicable to the case of $C_0 = \{\Phi\}$ and $C_1 = \mathrm{Conv}(\mathbb{U}(\mathcal{H}))$. An interesting fact is that without auxiliary systems, we cannot operationally distinguish between a unital quantum channel Φ and $\mathrm{Conv}(\mathbb{U}(\mathcal{H}))$ even when the former is not contained in the latter. To see this, let $\rho \in \mathrm{L}(\mathcal{H})$ be any density operator. Since Φ is a unital quantum channel, it is also a doubly stochastic map. Thus we have the majorization relation $\Phi(\rho) \prec \rho$ [22]. By another Theorem of Uhlmann [24], we know there exist a probability distribution $\{p_k\}$ and a set of unitary operators $\{U_k\}$ such that

$$\Phi(\rho) = \sum_k p_k U_k \rho U_k^\dagger.$$

So $\widetilde{D}(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H})); \rho) = 0$ for any density operator ρ from $\mathrm{L}(\mathcal{H})$. On the other hand, we have $D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H}))) > 0$ even when the input can only be chosen from $\mathrm{L}(\mathcal{H})$. This indicates that D and $\widetilde{D}$ are quite different when we do not use auxiliary systems.

IV. A LOWER BOUND FOR THE DISTANCE BETWEEN A QUANTUM CHANNEL AND THE CONVEX HULL OF UNITARY CHANNELS

It is generally difficult to decide whether a given unital quantum channel Φ is a mixture of unitary channels or not. One simple sufficient condition is that the Kraus operator space $K(\Phi)$ does not contain any unitary operator, i.e., $K(\Phi) \cap \mathbb{U}(\mathcal{H}) = \emptyset$. (Note that the Kraus operator space $K(\Phi) = \mathrm{span}\{E_k\}$ for a quantum channel $\Phi = \sum_k E_k \cdot E_k^\dagger$. If this is the case, we can actually obtain an analytical lower bound for the distance between Φ and $\mathrm{Conv}(\mathbb{U}(\mathcal{H}))$.

Lemma 2. *For any quantum channel $\Phi \in \mathrm{T}(\mathcal{H}_d)$ such that $K(\Phi) \cap \mathbb{U}(\mathcal{H}_d) = \emptyset$, we have*

$$D(\Phi, \mathrm{Conv}(\mathbb{U}(\mathcal{H}_d))) \geq \min_{L \in K(\Phi)} \frac{\mathrm{Tr}(|L| - I_d)^2}{d} = C_\Phi > 0.$$

Proof: Let $\Psi = \sum_{k=1}^N p_k \mathcal{U}_k$ with $\{p_k\}$ a finite probability distribution and $\mathcal{U}_k \in \mathbb{U}(\mathcal{H}_d)$. We need to show that

$$D(\Phi, \Psi) = D(\Phi, \sum_k p_k \mathcal{U}_k) \geq C_\Phi.$$

Note that

$$
\begin{aligned}
D(\Phi, \sum_k p_k \mathcal{U}_k) =& D(\Phi \otimes I_{\mathrm{L}(\mathcal{Z}_d)}, \sum_k p_k \mathcal{U}_k \otimes I_{\mathrm{L}(\mathcal{Z}_d)}) \\
\geq& D(\Phi \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega), \sum_k p_k \mathcal{U}_k \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega)),
\end{aligned}
$$

where $|\Omega\rangle = 1/\sqrt{d} \sum_{k=1}^d |k\rangle|k\rangle$ is a maximally entangled state on $\mathcal{H}_d \otimes \mathcal{Z}_d$. Now applying the inequality $D(\rho, \sigma) \geq 2(1 - F(\rho, \sigma))$, we have

$$
\begin{aligned}
D(\Phi, \Psi) \geq& 2(1 - F(\Phi \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega), \sum_k p_k \mathcal{U}_k \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega))) \\
=& 2(1 - \max_\psi |\langle \psi | \phi \rangle|),
\end{aligned}
$$

where $|\psi\rangle = \sum_k \sqrt{q_k} |\psi_k\rangle |k_\mathcal{K}\rangle$ ranges over all purifications of $\Phi \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega)$, $|\phi\rangle$ is a fixed purification of $\sum_k p_k \mathcal{U}_k \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega)$ given by

$$|\phi\rangle = \sum_k \sqrt{p_k}(U_k \otimes I_{\mathcal{Z}_d})|\Omega\rangle \otimes |k_\mathcal{K}\rangle,$$

$\{|k_\mathcal{K}\rangle\}$ is a fixed orthonormal basis for an auxiliary system $\mathcal{K}$, $\{q_k\}$ is a probability distribution, and $|\psi_k\rangle$ are unit vectors in $\mathcal{H}_d \otimes \mathcal{Z}_d$. An important observation here is that $|\psi_k\rangle$ is in the support of $\Phi \otimes I_{\mathrm{L}(\mathcal{Z}_d)}(\Omega)$ which is spanned by a set of vectors of the form $(E_j \otimes I_{\mathcal{Z}_d})|\Omega\rangle$, where we assume that $\Phi = \sum_j E_j \cdot E_j^\dagger$. Hence

$$|\psi_k\rangle = \sum_j \lambda_j (E_j \otimes I_{\mathcal{Z}_d})|\Omega\rangle$$

for some complex numbers λ_j, from which we readily deduce that

$$|\psi_k\rangle = (L_k \otimes I_{\mathcal{Z}_d})|\Omega\rangle,$$

where

$$L_k = \sum_j \lambda_j E_j \in K(\Phi).$$

Since $|\psi_k\rangle$ are unit vectors, $\mathrm{Tr}|L_k|^2 = \mathrm{Tr}(L_k^\dagger L_k) = d$. Thus we have

$$
\begin{aligned}
D(\Phi, \Psi) \geq& 2(1 - \max_{L_k, q_k} |\sum_k \sqrt{p_k q_k} \mathrm{Tr}(L_k^\dagger U_k)/d|\}) \\
\geq& 2(1 - \max\{\mathrm{Tr}(L_k^\dagger U_k)/d : L_k \in K(\Phi)\}) \\
\geq& 2(1 - \max\{\frac{1}{d}\mathrm{Tr}(L^\dagger U) : \mathrm{Tr}|L|^2 = d, U \in \mathbb{U}(\mathcal{H}_d)\}) \\
=& 2(1 - \frac{1}{d}\max\{\mathrm{Tr}|L| : \mathrm{Tr}|L|^2 = d, L \in K(\Phi)\}) \\
=& \min\{\frac{1}{d}\mathrm{Tr}(|L| - I_{\mathcal{H}_d})^2 : \mathrm{Tr}|L|^2 = d, L \in K(\Phi)\} \\
\geq& \inf\{\frac{1}{d}\mathrm{Tr}(|L| - I_{\mathcal{H}_d})^2 : L \in K(\Phi)\}.
\end{aligned}
$$

In the last step we have to use "inf" instead of "min" as the domain of L has been broadened from a compact set $\{L \in K(\Phi) : \text{Tr}|L|^2 = d\}$ to an unbounded set $K(\Phi)$. To finish the proof, we need to show that "inf" in the last line can be replaced by "min". First, notice that the RHS of the above equation is less than 2, and

$$\frac{\text{Tr}(|L| - I_{\mathcal{H}_d})^2}{d} \geq \frac{(\text{Tr}|L| - d)^2}{d^2}.$$

If $\text{Tr}|L| \geq (\sqrt{2} + 1)d$ then the right hand side (RHS) of the above equation is greater than 2. Thus

$$\inf_{L \in K(\Phi)} \frac{\text{Tr}(|L| - I_{\mathcal{H}_d})^2}{d} = \min_{\text{Tr}|L| \leq (1+\sqrt{2})d} \frac{\text{Tr}(|L| - I_{\mathcal{H}_d})^2}{d}.$$

As a final remark, we need show that $C_\Phi > 0$ under the assumption $K(\Phi) \cap U(\mathcal{H}_d) = \emptyset$. Otherwise, $C_\Phi = 0$ implies that there is some $\widetilde{L} \in K(\Phi)$ such that $|\widetilde{L}| = I_{\mathcal{H}_d}$. In other words, $\widetilde{L}$ is unitary, which is a contradiction. $\square$

Theorem 2. *Let $\Phi \in T(\mathcal{H}_d)$ be a quantum channel, and let $\Psi \in S(\mathcal{H}_m)$ be any Schur channel. Then*

$$D(\Psi \otimes \Phi, \text{Conv}(U(\mathcal{H}_m \otimes \mathcal{H}_d))) \geq C_\Phi.$$

Proof: The key observation here is that under the assumption Ψ is with diagonal Kraus operators. Thus any $L \in K(\Psi \otimes \Phi)$ can be decomposed as

$$L = \oplus_{k=1}^m L_k, \quad L_k \in K(\Phi).$$

Suppose now that $\widetilde{L} = \oplus_{k=1}^m \widetilde{L}_k$ achieves the minimum in $C_{\Psi \otimes \Phi}$. We have

$$
\begin{aligned}
C_{\Psi \otimes \Phi} &= \frac{1}{md} \text{Tr}(|\widetilde{L}| - I_{\mathcal{H}_m \otimes \mathcal{H}_d})^2 \\
&= \frac{1}{md} \text{Tr}(\oplus_{k=1}^m (|\widetilde{L}_k| - I_{\mathcal{H}_d}))^2 \\
&= \frac{1}{md} \text{Tr}(\oplus_{k=1}^m (|\widetilde{L}_k| - I_{\mathcal{H}_d})^2) \\
&= \frac{1}{md} \sum_{k=1}^m \text{Tr}(|\widetilde{L}_k| - I_{\mathcal{H}_d})^2 \\
&\geq C_\Phi,
\end{aligned}
$$

where we have employed the fact that

$$\text{Tr}(|\widetilde{L}_k| - I_{\mathcal{H}_d})^2 \geq dC_\Phi, \forall\, 1 \leq k \leq m.$$

Now the desired result follows from Lemma 2 directly. $\square$

As a direct corollary, we have the following

Corollary 1. *For any Schur channel $\Phi \in S(\mathcal{H}_d)$, if $K(\Phi) \cap U(\mathcal{H}_d) = \emptyset$, then*

$$D(\Phi^{\otimes n}, \text{Conv}(U(\mathcal{H}_d^{\otimes n}))) \geq C_\Phi > 0, \quad \forall n \geq 1.$$

V. BOUNDS ON THE DISTANCE BETWEEN A SCHUR CHANNEL AND THE CONVEX HULL OF UNITARY CHANNELS

The condition that the Kraus operator space $K(\Phi)$ of Φ does not contain any unitary operator is a very strong constraint. In most cases we may have that Φ is not a mixture of unitary channels but $K(\Phi)$ contains some unitary operator. Here we deal with this more general case but only for Schur channels. In this case we are able to show that up to a factor of $1/2$, any Schur channel can be approximated by a mixture of diagonal unitary channels.

Let us denote

$$\Lambda(\mathcal{H}_d) = S(\mathcal{H}_d) \cap \text{Conv}(U(\mathcal{H}_d)).$$

Intuitively, $\Lambda(\mathcal{H}_d)$ (or simply Λ_d) is the set of Schur channels that are also mixtures of diagonal unitary channels. So any $\Psi \in \Lambda_d$ can be written into the form $\Psi = \sum_k p_k U_k \cdot U_k^\dagger$, where U_k are $d \times d$ diagonal unitary matrices.

Theorem 3. *For given Schur channel $\Phi \in S_d$, we have*

$$\frac{1}{2} D(\Phi, \Lambda_d) \leq D(\Phi, \text{Conv}(U_d)) \leq D(\Phi, \Lambda_d). \quad (8)$$

Proof: The second inequality follows directly from $\Lambda_d \subset \text{Conv}(U_d)$. We will employ some standard arguments in operator algebras to prove the first inequality. Let $\Psi = \sum_k p_k \mathcal{U}_k \in \text{Conv}(U_d)$ such that

$$D(\Phi, \Psi) = D(\Phi, \text{Conv}(U_d)) = \delta.$$

We only need to prove that

$$D(\Phi, \Lambda_d) \leq 2\delta.$$

For any two diagonal unitary matrices U and V, let us introduce a map $J^{U,V} : T(\mathcal{H}_d) \to T(\mathcal{H}_d)$ as follows:

$$J^{U,V}(\Phi) = U^\dagger \Phi(U \cdot V) V^\dagger.$$

It is obvious that $J^{U,V}$ is an isometry over $T(\mathcal{H}_d)$ in the following sense:

$$D(J^{U,V}(\Phi_1), J^{U,V}(\Phi_2)) = D(\Phi_1, \Phi_2) \quad (9)$$

for any $\Phi_1, \Phi_2 \in T(\mathcal{H}_d)$.

Now we can further introduce a map $J : T(\mathcal{H}_d) \to T(\mathcal{H}_d)$ such that

$$J(\Phi) = \int_{\widehat{U}_d} \int_{\widehat{U}_d} J^{U,V}(\Phi) dU dV,$$

where both dU and dV are Haar measures over the diagonal unitary group $\widehat{U}_d$. The map J satisfies the following properties:

i). J is a contraction in the sense

$$D(J(\Phi_1), J(\Phi_2)) \leq D(\Phi_1, \Phi_2), \forall \Phi_1, \Phi_2 \in T(\mathcal{H}_d),$$

which is a simple consequence of the convexity of the diamond norm and Eq. (9).

ii). $J(\Phi) = \Phi$ for any Schur multiplier $\Phi \in T(\mathcal{H}_d)$. This is true simply due to the following observation

$$J^{U,V}(\Phi) = \Phi,$$

where $U, V \in \widehat{U}_d$ are diagonal unitary matrices.

iii). $J(\Phi)$ is a Schur multiplier for any $\Phi \in T(\mathcal{H}_d)$. In particular, $J(\Phi)$ is CP whenever Φ is CP. To see that, by a direct calculation, we find that for $\Phi = \sum_k E_k \cdot F_k^\dagger$,

$$J(\Phi) = \sum E_k' \cdot F_k'^\dagger, \ E_k' = \mathrm{diag}(E_k), F_k' = \mathrm{diag}(F_k).$$

Clearly, $J(\Phi)$ is a Schur multiplier. When Φ is CP, we can choose

$$E_k' = F_k' = \mathrm{diag}(E_k) = \mathrm{diag}(F_k),$$

thus $J(\Phi)$ is CP.

Now we can compute that

$$\Psi' = J(\Psi) = \sum_k p_k J(\mathcal{U}_k) = \sum_k p_k A_k \cdot A_k^\dagger,$$

where $A_k = \mathrm{diag}(U_k)$. We have

$$D(\Phi, \Lambda_d) \leq D(\Phi, \Psi') + D(\Psi', \Lambda_d). \qquad (10)$$

The first term in the RHS of Eq. (10) satisfies

$$D(\Phi, \Psi') = D(J(\Phi), J(\Psi)) \leq D(\Phi, \Psi) = \delta,$$

where we have employed the contraction property of J, item i) above.

It remains to show that the second term in the RHS of Eq. (10) fulfills

$$D(\Psi', \Lambda_d) \leq \delta.$$

Our strategy is to choose $\Psi'' \in T(\mathcal{H}_d)$ such that

$$D(\Psi', \Psi'') \leq \delta, \ \Psi'' \in \Lambda_d \qquad (11)$$

Then

$$D(\Psi', \Lambda_d) \leq D(\Psi', \Psi'') \leq \delta.$$

The rest of the proof devotes to finding such Ψ''. Notice that each Kraus operator A_k of Ψ' is a diagonal contraction. Applying a well-known fact in linear algebra, we can choose two diagonal unitary matrices V_k, W_k such that

$$A_k = \frac{1}{2}(V_k + W_k).$$

Now define

$$\Psi'' = \sum_k \frac{p_k}{2}(V_k \cdot V_k^\dagger + W_k \cdot W_k^\dagger).$$

Clearly $\Psi'' \in \Lambda_d$. We will show that Ψ'' satisfies Eq. (11). First, we find that $\Psi'' - \Psi'$ is a CP map. By a direct calculation, we have

$$\Psi'' - \Psi' = \sum_k \frac{p_k}{4}(W_k - V_k) \cdot (W_k - V_k)^\dagger.$$

Thus

$$(\Psi'' - \Psi')^\dagger = \Psi''^\dagger - \Psi'^\dagger$$

is also a CP map.

Now employing essentially the same techniques first introduced by Haagerup and Musat in [13], we have

$$\begin{aligned}
||\Psi'' - \Psi'||_\diamond &= ||\Psi''^\dagger - \Psi'^\dagger||_{\mathrm{cb}} \\
&= ||\Psi''^\dagger - \Psi'^\dagger||_\infty \\
&= ||(\Psi''^\dagger - \Psi'^\dagger)(I_{\mathcal{H}_d})||_\infty \\
&= ||I_{\mathcal{H}_d} - \Psi'^\dagger(I_{\mathcal{H}_d})||_\infty \\
&= ||\Phi^\dagger(I_{\mathcal{H}_d}) - \Psi'^\dagger(I_{\mathcal{H}_d})||_\infty \\
&\leq ||\Phi^\dagger - \Psi'^\dagger||_\infty \\
&= ||\Phi - \Psi'||_1 \\
&\leq ||\Phi - \Psi'||_\diamond.
\end{aligned}$$

That means

$$D(\Psi', \Psi'') \leq D(\Phi, \Psi') \leq \delta.$$

$\square$

It seems quite likely that in Eq. (8) the first inequality should be strict and the second one should be an equality. However, this is still an unsettled issue.

Theorem 4. *For given Schur channel $\Phi \in S_d$ and arbitrary $\Psi \in S_m$, we have*

$$D(\Psi \otimes \Phi, \Lambda_{m \otimes d}) \geq D(\Phi, \Lambda_d), \forall \Psi \in S_m. \qquad (12)$$

Here $m \otimes d$ is a shorthand for $\mathcal{H}_m \otimes \mathcal{H}_d$.

Proof: To show Eq. (12), we first choose $\Psi' = \sum_k p_k U_k \cdot U_k^\dagger \in \Lambda_{m \otimes d}$ such that

$$D(\Psi \otimes \Phi, \Lambda_{m \otimes d}) = D(\Psi \otimes \Phi, \Psi'). \qquad (13)$$

Notice that any diagonal unitary matrix U_k can be written into the following form:

$$U_k = \sum_{j=1}^m |j\rangle\langle j| \otimes U_j^{(k)},$$

where $U_j^{(k)} \in \widehat{U}_d$ are all diagonal unitary matrices. That implies

$$\Psi'' = \langle 1|\Psi|1\rangle = \sum_k p_k \mathcal{U}_1^{(k)} \in \Lambda_d,$$

where $\mathcal{U}_1^{(k)}$ is the unitary channel corresponding to $U_1^{(k)}$. Intuitively, the compressed version Ψ'' of Ψ' remains a mixture of diagonal unitary channels. Now we have

$$
\begin{aligned}
D(\Psi \otimes \Phi, \Psi') &\geq \sup_{||X||_1 \leq 1} ||(\Psi \otimes \Phi - \Psi')(|1\rangle\langle 1| \otimes X)||_1 \\
&= \sup_{||X||_1 \leq 1} |||1\rangle\langle 1| \otimes (\Phi - \Psi'')(X)||_1 \\
&= ||\Phi - \Psi''||_1 \\
&= ||\Phi - \Psi''||_\diamond \\
&\geq D(\Phi, \Lambda_d),
\end{aligned}
$$

where we have used the fact that $\Psi(|1\rangle\langle 1|) = |1\rangle\langle 1|$ and $\Psi'' = \langle 1|\Psi'|1\rangle \in \Lambda_d$.

Combining the above equation with Eq. (13), we have proven Eq. (12). $\square$

A somewhat interesting fact is that the above two results together can be used to derive some results first obtained by Haagerup and Musat in Ref. [13], which are applicable to the AQBC.

Theorem 5. *(Haagerup and Musat [13]) For given Schur channel $\Phi \in S_d$ and arbitrary $\Psi \in S_m$, we have*

$$
D(\Psi \otimes \Phi, \mathrm{Conv}(\mathbb{U}_{m \otimes d})) \geq \frac{1}{2} D(\Phi, \mathrm{Conv}(\mathbb{U}_d)). \quad (14)
$$

Proof: Actually Eq. (14) is a quite straightforward application of the above two theorems. First notice that $\Psi \otimes \Phi \in S_{m \otimes d}$. Applying Theorem 3 to $\Psi \otimes \Phi$, we have

$$
D(\Psi \otimes \Phi, \mathrm{Conv}(\mathbb{U}_{m \otimes d})) \geq \frac{1}{2} D(\Psi \otimes \Phi, \Lambda_{m \otimes d}).
$$

On the other hand, it is obvious that

$$
D(\Phi, \Lambda_d) \geq D(\Phi, \mathrm{Conv}(\mathbb{U}_d)).
$$

Thus the left thing is to show

$$
D(\Psi \otimes \Phi, \Lambda_{m \otimes d}) \geq D(\Phi, \Lambda_d),
$$

and this is exactly the content of Theorem 4. $\square$

Corollary 2. *(Haagerup and Musat [13]) Let $\Phi \in S_d$ be a Schur channel that does not satisfy the quantum Birkhoff property, that is, $\Phi \notin \mathrm{Conv}(\mathbb{U}_d)$. Then Φ does not satisfy the asymptotic quantum Birkhoff property, and*

$$
D(\Phi^{\otimes n}, \mathrm{Conv}(\mathbb{U}_{d^{\otimes n}})) \geq \frac{1}{2} D(\Phi, \mathrm{Conv}(\mathbb{U}_d)).
$$

VI. SOME EXPLICIT COUNTEREXAMPLES TO THE ASYMPTOTIC QUANTUM BIRKHOFF CONJECTURE

Our results in Section IV enable us to construct counterexamples to AQBC easily. Our basic strategy is to construct Schur channel Φ satisfying $K(\Phi) \cap U(\mathcal{H}) = \emptyset$. Then the statement that Φ is a counterexample to AQBC follows directly from Corollary 1.

Example 1. Our first example is chosen from Ref. [7] (Section 4.3). $\Phi = E_1 \cdot E_1^\dagger + E_2 \cdot E_2^\dagger$, where

$$
E_1 = \mathrm{Diag}(1, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), \; E_2 = \mathrm{Diag}(0, 1, \frac{1}{\sqrt{2}}, -\frac{i}{\sqrt{2}}).
$$

Clearly none of E_1 and E_2 is unitary. We now show that there is no unitary in $K(\Phi)$. By contradiction, assume that for some complex numbers λ and μ we have that $\lambda E_1 + \mu E_2$ is unitary. Then

$$
(\lambda E_1 + \mu E_2)^\dagger (\lambda E_1 + \mu E_2) = I_4,
$$

from which we obtain

$$
|\lambda|^2 = 1, |\mu|^2 = 1, \frac{1}{2}(|\lambda|^2 + |\mu|^2) = 1, \frac{1}{2}(|\lambda|^2 - |\mu|^2) = 1.
$$

Clearly, there is no λ and μ satisfying all the above equations. Thus we have $K(\Phi) \cap U(\mathcal{H}_4) = \emptyset$. It follows from Corollary 1 that Φ is a counterexample to AQBC. One can readily verify that the set $\{E_1^\dagger E_1, E_1^\dagger E_2, E_2^\dagger E_1, E_2^\dagger E_2\}$ is linearly independent. Thus it follows from Corollary 2.3 of Ref. [11] that Φ is a non-factorizable map. $\square$

Example 2. Our second example is taken from Ref. [11] (Example 3.3). $\Phi = \sum_{k=1}^3 E_k \cdot E_k^\dagger$, where

$$
E_1 = \mathrm{Diag}(1, \frac{1}{\sqrt{5}} I_5), E_2 = \mathrm{Diag}(0, \sqrt{\frac{2}{5}} Z_5), E_3 = E_2^\dagger,
$$

where $Z_5 = \mathrm{Diag}(1, \frac{2\pi i}{5}, \frac{4\pi i}{5}, \frac{6\pi i}{5}, \frac{8\pi i}{5})$ satisfying $Z_5^5 = I_5$. In the following we directly write I and Z for I_5 and Z_5, respectively.

As shown in [11], one can choose a set of Hermitian Kraus operators F_1, F_2, F_3 such that

$$
F_1 = E_1, \; F_2 = \frac{1}{2}(E_2 + E_3), \; F_3 = \frac{1}{2i}(E_2 - E_3).
$$

It is easy to see that $\Phi = \sum_{k=1}^3 F_k \cdot F_k^\dagger$. By Corollary 2.5 of Ref. [11], Φ is a factorizable map.

Now we show that $K(\Phi) \cap U(\mathcal{H}_6) = \emptyset$. Again by contradiction, assume there are complex numbers $\lambda_1, \lambda_2, \lambda_3$ such that $\lambda_1 E_1 + \lambda_2 E_2 + \lambda_3 E_3$ is a unitary. In other words, $\mathrm{Diag}(\lambda_1, \sqrt{\frac{1}{5}}\lambda_1 I + \sqrt{\frac{2}{5}}\lambda_2 Z + \sqrt{\frac{2}{5}}\lambda_3 Z^{-1})$ is a unitary. This is equivalent to

$$
|\lambda_1|^2 = 1, \; |\lambda_1 I + \sqrt{2}\lambda_2 Z + \sqrt{2}\lambda_3 Z^{-1}| = \sqrt{5} I.
$$

For simplicity, we may assume $\lambda_1 = 1$, $a = \sqrt{2}\lambda_2$, and $b = \sqrt{2}\lambda_3$. Then we can rewrite the above equation as follows:

$$
(|a|^2 + |b|^2 - 4)I + (a + b^*)Z + (a^* + b)Z^{-1} + ab^* Z^2 + ab^* Z^{-2} = 0.
$$

Employing the fact that $\{I, Z, Z^{-1}, Z^2, Z^{-2}\}$ are linearly independent, we have

$$
|a|^2 + |b|^2 = 4, a + b^* = 0, ab^* = 0.
$$

Clearly, there are no a and b satisfying all the above equations.

Hence Corollary 1 is applicable. This gives us a factorizable map which is also a counterexample to AQBC. This fact has been pointed out in the published version of Ref. [11], and was derived by the result in Ref. [13]. $\square$

As a matter of fact, all counterexamples to AQBC presented above are simply the counterexamples to QBC. It would be quite interesting to know for what kind of unital channels Φ these two properties are different, i.e., Φ is a counterexample to QBC, but fulfills AQBC. A systematic way to construct unital channels that violate QBC has been proposed by Bravyi and Smolin using the idea of unextendible maximally entangled bases [25]. All unital channels Φ constructed in this way will automatically satisfy the condition $K(\Phi) \cap U(\mathcal{H}) = \emptyset$. However, it remains a formidable task to verify whether these unital channels fulfill or violate the AQBC. A preliminary step towards this goal is to invent some tractable upper bounds for the distance between a unital channel and the convex hull of unitary channels.

VII. A CONNECTION TO GROTHENDIECK'S INEQUALITY

It is well known that Grothendieck's inequality (GI) in the metric theory of tensor products is closely related to Bell's inequality in quantum information theory [26]. As an interesting application of the results in Sections IV and V, we explain here that GI has intimate links with AQBC too. Let us first recall the equivalent formulation of GI in terms of Schur multipliers. Let

$$\mathbb{S}_d = \{\Phi_S : S \in \mathrm{L}(\mathcal{H}_d), \|\Phi_S\|_1 \leq 1\}.$$

Namely, $\mathbb{S}_d$ is the unit ball of the space of all Schur multipliers on $\mathrm{L}(\mathcal{H}_d)$ with respect to the trace norm or any of the three other norms considered before (see Proposition 3). It is well known that $\Phi_S \in \mathbb{S}_d$ if and only if there exists a Hilbert space $\mathcal{K}$ and vectors $|\xi_1\rangle, ..., |\xi_d\rangle, |\eta_1\rangle, ..., |\eta_d\rangle$ in the unit ball of $\mathcal{K}$ such that

$$s_{kj} = \langle \xi_k | \eta_j \rangle. \tag{15}$$

Here we have assumed that $S = [s_{kj}]$ with respect to a fixed orthonormal basis of $\mathcal{H}_d$. Note that $\mathcal{K}$ can be chosen to be finite dimensional. This representation of Φ_S is to be compared with i) of Proposition 2. Indeed, developing ξ_k and η_j in an orthonormal basis of $\mathcal{K}$, we recover the representation of Φ_S given by Proposition 2.

The case where $\dim \mathcal{K} = 1$ is of particular interest. The corresponding set of Schur multipliers is denoted by $\mathbb{D}_d$, that is, $\mathbb{D}_d$ is the set of all Schur multipliers Φ_S of the form

$$s_{kj} = \alpha_k^* \beta_j, \quad 1 \leq k, j \leq d, \tag{16}$$

where α_k, β_j are complex numbers such that

$$\max_k |\alpha_k| \leq 1, \quad \max_j |\beta_j| \leq 1.$$

It is clear that

$$\mathrm{Conv}(\mathbb{D}_d) \subseteq \mathbb{S}_d.$$

GI asserts that the converse inclusion also holds true up to a universal constant:

Grothendieck's inequality. *There exists a universal constant K such that for all $d \geq 1$*

$$\mathbb{S}_d \subseteq K\mathrm{Conv}(\mathbb{D}_d). \tag{17}$$

The smallest constant K is called Grothendieck's constant, denoted by K_G. The exact value of K_G is still unknown. But it is well known that $1 < K_G \leq 1.4049$. (Note here all scalars are assumed to be complex numbers). What is important for us is the fact that $K_G > 1$. We refer to Chapter 5 of Ref. [27] and page 19 of Ref. [28] for more information.

It is easy to see that a Schur multiplier Φ_S is positive if and only if we can choose ξ_k and η_k in (15) such that $\xi_k = \eta_k$ for all $1 \leq k \leq d$. Let $\mathbb{S}_d^+$ denote the positive part of $\mathbb{S}_d$. Accordingly, let $\mathbb{D}_d^+$ denote the positive part of $\mathbb{D}_d$. Namely, $\mathbb{D}_d^+$ is the set of all Schur multipliers of the form (16) with $\alpha_k = \beta_k$. We again have obviously

$$\mathrm{Conv}(\mathbb{D}_d^+) \subseteq \mathbb{S}_d^+.$$

Surprisingly, this time the converse inclusion does not hold up to a universal constant. More precisely, let K_d^+ denote the least constant such that

$$\mathbb{S}_d^+ \subseteq K_d^+ \mathrm{Conv}(\mathbb{D}_d^+).$$

Then we have the following result of Kashin and Szarek from Ref. [29] (see also the lemma on the page 17 of Ref. [28]).

Proposition 4. *There exist two positive constants α and β such that $\alpha \log d \leq K_d^+ \leq \beta \log d$ for all $d > 1$.*

The Schur multipliers we are interested in are Schur channels (unital positive Schur multipliers). Recall that the set of all Schur channels on $\mathrm{L}(\mathcal{H}_d)$ have been denoted by S_d in the previous sections. This set could be also denoted by $\mathbb{S}_{d,1}^+$ in the current notational system (1 being for "unital"). Accordingly, Λ_d is the subset of $\mathrm{Conv}(\mathbb{D}_d^+)$ consisting of Schur channels. As shown in sections IV and V, to disprove AQBC is equivalent to showing that the obvious inclusion

$$\Lambda_d \subseteq S_d \tag{18}$$

is strict for some d. We now show that this inclusion is strict for large d in the spirit of Proposition 4. To this end let $K_{d,1}^+$ denote the least constant K such that

$$S_d \subseteq K\mathrm{Conv}(\mathbb{D}_d^+).$$

Then inclusion (18) is strict if and only if $K_{d,1}^+ > 1$.

Proposition 5. $K_{d,1}^+ = K_d^+$ *for all* d.

Proof: It is clear that $K_{d,1}^+ \leq K_d^+$. To prove the converse inequality, let $\Phi_S \in \mathbb{S}_d^+$. Then there exist a Hilbert space $\mathcal{K}$ and vectors $\xi_1, ..., \xi_d$ in the unit ball of $\mathcal{K}$ such that

$$s_{kj} = \langle \xi_k | \xi_j \rangle.$$

Without loss of generality, we can assume that $\xi_k \neq 0$ for all k. Let

$$\eta_k = \frac{\xi_k}{\|\xi_k\|} \quad \text{and} \quad t_{kj} = \langle \eta_k | \eta_j \rangle.$$

Then $\Phi_T \in S_d \subseteq K_{d,1}^+ \text{Conv}(\mathbb{D}_d^+)$, so there exist complex numbers $\alpha_{k,i}$ of modulus not greater than 1 and positive numbers λ_i such that

$$t_{kj} = K_{d,1}^+ \sum_i \lambda_i \alpha_{k,i}^* \alpha_{j,i} \quad \text{and} \quad \sum_i \lambda_i = 1.$$

It follows that

$$s_{kj} = K_{d,1}^+ \sum_i \lambda_i (\|\xi_k\| \alpha_{k,i}^*)(\|\xi_j\| \alpha_{j,i}) \in K_{d,1}^+ \text{Conv}(\mathbb{D}_d^+).$$

We then deduce $K_d^+ \leq K_{d,1}^+$. $\qquad\square$

Consequently, $K_{d,1}^+ \approx \log d$ as $d \to \infty$. This implies that inclusion (18) is strict for large d. Therefore, for any sufficiently large d, there exists a Schur channel Φ over $L(\mathcal{H}_d)$ that is not a mixture of unitary Schur channels. As direct consequences of the results in section V, such a Schur channel must violate AQBC.

By Ref. [7] the first integer d for which $K_{d,1}^+ > 1$ is $d = 4$. Thus $K_d^+ = K_{d,1}^+ = 1$ if and only if $d \leq 3$. This means that inclusion (18) is an equality for $d \leq 3$ and becomes strict for $d \geq 4$.

On the other hand, denote K_d the least constant K in (17) for a fixed d. Note that $K_G = \sup_d K_d$. It was proved independently by Davie and the third named author that $K_2 = 1$ (see the remark at the end of Chapter 5 of Ref. [27]). It seems, however, that the first integer d for which $K_d > 1$ is still unknown. This problem is related to the characterization of the extreme points of $\mathbb{S}_d$. Indeed, $K_d > 1$ is equivalent to the existence of extreme points Φ_S of $\mathbb{S}_d$ that are not of the form $S = [\alpha_k^* \beta_j]$ for some complex numbers α_k and β_k with $|\alpha_k| = |\beta_k| = 1$ for all k. Thus such extreme points exist for large d. It would be interesting to characterize the extreme points of $\mathbb{S}_d$ in the spirit of Refs. [2] and [7].

[1] K. Kraus, States, effects, and operations, Springer-Verlag, Berlin, 1983.

[2] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Alg. Appl. **10**, 285 (1975).

[3] M. Gregoratti and R. F. Werner, Quantum lost and found, J. Mod. Opt. **50**, 915 (2002).

[4] J.A. Smolin, F. Verstraete, and A. Winter, Entanglement of assistance and multipartite state distillation, Phys. Rev. A **72**, 052317 (2005).

[5] B. Kummerer, Construction and strucure of Markov dillations on W^*-algebra, Habilitationsschrift, Tuibingen, 1986.

[6] B. Kummerer and H. Maasen, The essentially commutative dilations of dynamical semigroups on M_n, Commun. Math. Phys. **109**, 1-22 (1987).

[7] L.J. Landau and R.F. Streater, On Birkhoff's theorem for doubly stochastic completely positive maps on matrix algebra, Lin. Alg. Applic. **193**, 107-127 (1993).

[8] R. Werner, Open problems in quantum information theory,URL: http://qig.itp.uni-hannover.de/qiproblems/30.

[9] V. Paulsen, Completely bounded maps and operator algebras, Combridge University Press, 2002.

[10] C. B. Mendl and M. M. Wolf, Unital quantum channels - convex structure and revivals of Birkhoff's theorem, Commun. Math. Phys. **289**, 1057 (2009).

[11] U. Haagerup and M. Musat, Factorization and dilation problems for completely positive maps on von Neumann algebras, Commun. Math. Phys. **303**, 555 (2011). Ealier arXiv version: http://arxiv.org/abs/1009.0778.

[12] D. Ostrev, A. Oza, and P. Shor, The structure of unital maps and the asymptotic quan-

tum Birkhoff conjecture, in preparation. See P. Shor's Steklov Mathematical Institute Seminar: http://www.mathnet.ru/php/seminars.phtml?option_lang=eng&presentid=1208.

[13] U. Haagerup and M. Musat, Factorizable completely positive maps and the Connes embedding problem, in preparation.

[14] J. Watrous, Notes on super-opertator norms induced by Schatten norms, http://arxiv.org/abs/0411077.

[15] J. Watrous, Semidefinite programs for completely bounded norms, http://arxiv.org/abs/0901.4709.

[16] C. W. Helstrom, Quantum detection and estimation theory, New York, Academic Press, 1976.

[17] M. Sion, On general minimax theorems, Pac. J. Math. **8**, 171 (1958).

[18] G. Gutoski and J. Watrous, Quantum interactive proofs with competing provers, http://arxiv.org/abs/cs/0412102.

[19] R. Jain, Distinguishing sets of quantum states, http://arxiv.org/abs/quant-ph/0506205.

[20] The possibility of such a proof was communicated to us by an anonymous referee of QIP'2012 in his/her review report and by Watrous in private conversation with the first named author (N. Yu). We would like to thank them for their helpful and generous comments.

[21] This follows from a simple but useful fact in matrix analysis: For any $C^\dagger C \le A^\dagger A$, there is $0 \le P \le I$ such that $C^\dagger C = A^\dagger P A$.

[22] Let x and y be two real vectors with the same dimension. we say that x is majorized by y, or $x \prec y$, if there is a doubly stochastic matrix D such that $x = Dy$. For two Hermitian matrices A and B, $A \prec B$ means $\lambda(A) \prec \lambda(B)$, where $\lambda(A)$ is the spectral vector of A, i.e., the real vector formed by all eigenvalues of A (counting the mutilicity of eigenvalues).

[23] G. Gutoski, On a measure of distance for quantum strategies, http://arxiv.org/abs/quant-ph/1008.4636.

[24] A. Uhlmann, Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory, Commun. Math. Phys. **54**, 21 (1977).

[25] S. Bravyi and J. A. Smolin, Unextendible maximally entangled bases, Phys. Rev. A **84**, 042306 (2011).

[26] B.S. Tsirelson, Quantum analogues of the Bell inequalities. The case of two spatially separated domains. Journal of Soviet Mathematics **36**, 557 - 570 (1987).

[27] G. Pisier, Similarity problems and completely bounded maps, Lecture Notes in Mathematics, **1618**, 2001.

[28] G. Pisier, Grothendieck's Theorem, past and present, http://arxiv.org/abs/1101.4195.

[29] B. Kashin and S. Szarek, On the Gram matrices of systems of uniformly bounded functions (Russian), Tr. Mat. Inst. Steklova 243 (2003), Funkts. Prostran., Priblizh., Differ. Uravn., 237 - 243; translation in Proc. Steklov Inst. Math. **243** (4), 227 (2003).